Noname manuscript No.
(will be inserted by the editor)

Vertex-Frequency Energy Distributions

Ljubiša Stanković, Miloš Daković, and Ervin Sejdić

Received: date / Accepted: date

Abstract Vertex-varying spectral content on graphs challenge the assumption of vertex invariance, and require vertex-frequency representations to adequately analyze them. The localization window in graph Fourier transform plays a crucial role in this analysis. An analysis of the window functions is presented. The corresponding spectrograms are considered from the energy condition point of view as well. Like in time-frequency analysis, the distribution of signal energy as a function of the vertex and spectral indices is an alternative way to approach vertex-frequency analysis. After an introduction to the second part of this chapter, a local smoothness definition, a definition of an ideal form of the vertex-frequency energy distributions, and two energy forms of the vertex-frequency representations are given. A graph form of the Rihaczek distribution is used as the basic distribution to define a class of reduced interference vertex-frequency energy distributions. These distributions reduce cross-terms effects and satisfy graph signal marginal properties. The theory is illustrated through examples.

1 Introduction

Graph signal processing is a challenging, but rapidly developing, topic. Many real world signals can be considered as graph signals, i.e., signals defined on a graph. Basic and advanced graph signal processing techniques are presented in [1–6]. Some of its applications in biomedical systems [7, 8] and big data analysis [1] are the best examples of its real-world potential.

L. Stanković and M. Daković
University of Montenegro, Podgorica, Montenegro
E-mail: ljubisa@ac.me, milos@ac.me

E. Sejdić
University of Pittsburg, Pittsburg, PA, USA
E-mail: esejdic@pitt.edu

In the case of large signals (graphs), we may not be interested in the analysis of the entire signal, but rather may be interested in its local behavior. A localized signal behavior can be examined via window functions. An exemplary analysis is signal averaging in a local neighborhood. This kind of processing corresponds to low-pass filtering in the classical time-domain signal analysis. Another example could be a classical time-frequency analysis [9–11], where we consider a local signal spectrum. In both examples, window functions are used in order to perform signal localization in time. Window functions are often symmetric, with a single maximum value at a considered time instant. Window functions can be easily shifted in time in order to analyze a signal's behavior at arbitrary time instants.

This concept of signal localization by using window functions can be extended to signals defined on graphs [12–16]. The extension is not straightforward since a simple operation like time shifting cannot be easily defined in a graph signal domain. Several solution approaches for this problem are defined.

A common approach is to utilize the signal spectrum to obtain the window functions for each graph vertex [6]. Another possibility is to define a window support as a local neighborhood for each vertex [16]. The localization window is defined by a set of vertices that contain the current vertex n and all vertices that are close to the vertex n. As in the classical signal analysis, a window should be narrow enough in order to provide good localization of the signal properties but wide enough to produce high resolution.

In this chapter, we will focus on the localized vertex spectrum of a graph signal. The basic concepts and localization methods are analyzed in Section 2. Vertex-frequency energy distributions are defined in Section 3, while their extension to the reduced interference vertex-frequency distributions is presented in Section 4.

2 Localized Graph Fourier Transform

2.1 Graph Fourier Transform

Consider a weighted undirected simple graph with N vertices, where edge weights are denoted by $w_{nm} > 0$ for an edge that connects a vertex n with a vertex m, and $w_{nm} = 0$ if vertices n and m are not connected with an edge. Edge weights are represented in a matrix form as a weight matrix $\mathbf{W}$ whose elements are w_{nm}. The diagonal matrix elements of $\mathbf{W}$ are zeros. The weighting matrix $\mathbf{W}$ is a symmetric matrix since the considered graph is undirected.

The signal $x(n)$, defined at each graph vertex n, is called the graph signal. Signal samples $x(n)$ can be arranged in a $N\times 1$ vector $\mathbf{x} = [x(1), x(2), \ldots, x(N)]^T$.

The graph Laplacian is defined as

$$\mathbf{L} = \mathbf{D} - \mathbf{W}, \tag{1}$$

where $\mathbf{D}$ is a diagonal degree matrix with $d_{nn} = \sum_{m=1}^{N} w_{nm}$ on the main diagonal. The eigenvalue decomposition of the Laplacian matrix reads

$$\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T, \tag{2}$$

where $\mathbf{U}$ is a matrix of eigenvectors (eigenvector $\mathbf{u}_k$ is the kth column of the matrix $\mathbf{U}$), and $\mathbf{\Lambda}$ is a diagonal matrix with eigenvalues λ_k on the main diagonal. Only simple eigenvalues of multiplicity one are assumed.

The spectrum of a graph signal $\mathbf{x}$ (the graph discrete Fourier transform GDFT) is defined as

$$\mathbf{X} = \text{GDFT}\{\mathbf{x}\} = \mathbf{U}^T\mathbf{x}, \tag{3}$$

where the vector $\mathbf{X}$ contains spectral coefficients associated to the kth eigenvalue and the corresponding eigenvector

$$X(k) = \mathbf{u}_k^T\mathbf{x} = \sum_{n=1}^{N} x(n)u_k(n). \tag{4}$$

The inverse transformation is obtained as $\mathbf{x} = \mathbf{U}\mathbf{X}$, with

$$x(n) = \sum_{k=1}^{N} X(k)u_k(n). \tag{5}$$

Here we will assume that the eigenvalues are of the multiplicity one. Approaches that extend GDFT to directed graphs and graphs with repeated eigenvalues may be found in [17–19].

2.2 Definition of the Localized Vertex Spectrum

The localized vertex spectrum (LVS) on a graph is an extension of the localized time (short time) Fourier transform (STFT). It can be calculated as the spectrum of a signal $x(n)$ multiplied by an appropriate localization window function $h_m(n)$

$$S(m,k) = \sum_{n=1}^{N} x(n)h_m(n)\ u_k(n). \tag{6}$$

It is assumed that the window function $h_m(n)$ should be such that it localizes the signal content around the vertex m. The local vertex spectrum in a matrix notation will be denoted as $\mathbf{S}$. Its elements are $S(m,k)$. The columns of $\mathbf{S}$ are

$$\mathbf{s}_m = \text{GDFT}\{x(n)h_m(n)\} = \mathbf{U}^T\mathbf{x}_{h_m},$$

where $\mathbf{x}_{h_m}$ is the vector whose elements $x(n)h_m(n)$ are equal to the signal samples multiplied by the window function centered at the vertex m.

For $h_m(n) = 1$, the localized vertex spectrum is equal to the standard spectrum $S(m,k) = X(k)$ for each m, i.e., no vertex localization is performed. If $h_m(m) = 1$ and $h_m(n) = 0$ for $n \neq m$, the localized vertex spectrum is

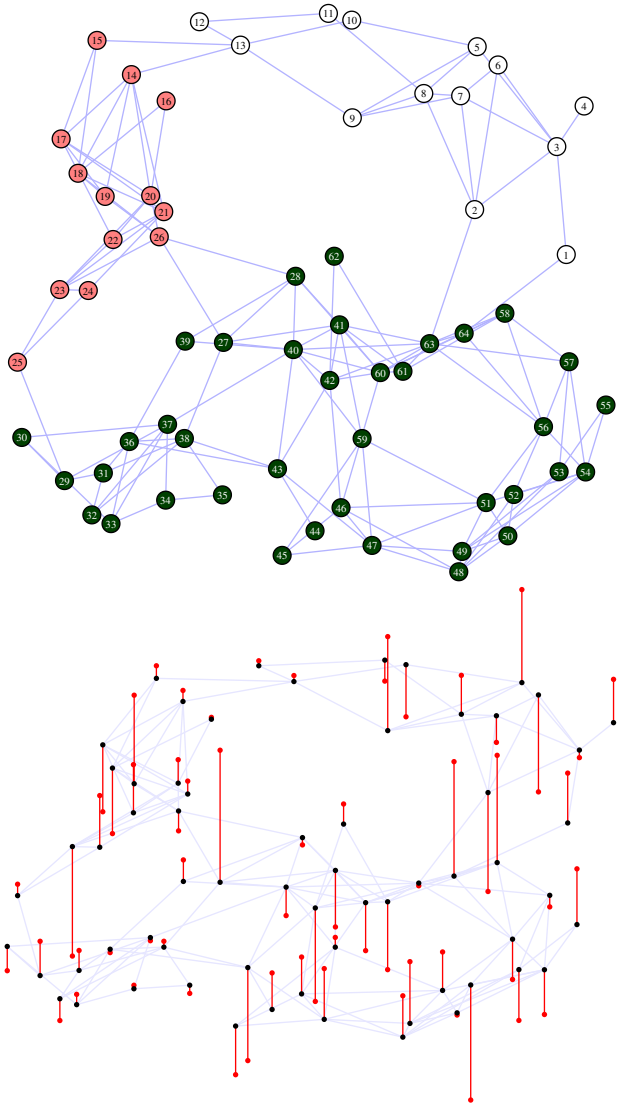

Fig. 1 Graph and signal on the graph. The signal is composed of three components. Component supports are presented with different vertex colors.

equal to the signal $S(m,k) = x(m)$ for each k and we do not have any spectral resolution.

Two methods for defining graph localization window functions $h_m(n)$ will be given. First we will present an analysis using the windows $h_m(n)$ defined in the vertex domain. In the next subsection, we will show how to create window functions $h_m(n)$ in the spectral domain.

2.3 Windows Defined Using the Vertex Neighborhood

The window $h_m(n)$ localized at vertex m can be defined using vertex neighborhood. The distance d_{mn} is equal to the length of the shortest walk from vertex m to vertex n. Note that d_{mn} are integers. Then the window function can be defined as

$$h_m(n) = g(d_{mn}),$$

where $g(d)$ corresponds to the basic window function in classical signal processing. For example, the Hann window can be used as

$$h_m(n) = \frac{1}{2}(1 + \cos(\pi d_{mn}/D)), \text{ for } 0 \leq d_{mn} < D,$$

where D is the assumed window width.

Window functions located at each vertex can be calculated in a matrix form. Vertices whose distance is $d_{mn} = 1$ are defined with an adjacency matrix $\mathbf{A}_1 = \mathbf{A}$. The adjacency matrix $\mathbf{A}$ is obtained from the weighting matrix $\mathbf{W}$, using the elements 1 if the vertices are connected and 0 if there is not an edge between the considered vertices. Vertices whose distance is $d_{mn} = 2$ are defined by the following matrix

$$\mathbf{A}_2 = (\mathbf{A} \odot \mathbf{A}_1) \circ (\mathbf{1} - \mathbf{A}_1) \circ (\mathbf{1} - \mathbf{I})$$

where $\odot$ is the logical (Boolean) matrix product, $\circ$ is the Hadamard (element-by-element) product, and $\mathbf{1}$ is a matrix with all ones. Matrix $\mathbf{A} \odot \mathbf{A}_1$ gives the information about all vertices that are connected with walks of length 2 and lower. Element-by-element multiplication by matrix $\mathbf{1} - \mathbf{A}_1$ removes the vertices connected with walks of length 1, while the multiplication by $\mathbf{1} - \mathbf{I}$ removes the diagonal elements.

For $d_{mn} = d \geq 2$ we have a recursive relation for the matrix that will give the information about the vertices at distance d

$$\mathbf{A}_d = (\mathbf{A} \odot \mathbf{A}_{d-1}) \circ (\mathbf{1} - \mathbf{A}_{d-1}) \circ (\mathbf{1} - \mathbf{I}).$$

The window matrix is then formed as

$$\mathbf{P}_D = g(0)\mathbf{I} + g(1)\mathbf{A}_1 + \cdots + g(D-1)\mathbf{A}_{D-1}.$$

The window weighted signal is formed using this matrix as

$$x_m(n) = h_m(n)x(n) = P_D(n,m)x(n).$$

The localized vertex-frequency representation is

$$S(m,k) = \sum_{n=1}^{N} x(n)h_m(n)\ u_k(n) = \sum_{n=1}^{N} x(n)P_D(n,m)\ u_k(n). \tag{7}$$

In matrix form, the above relation reads

$$\mathbf{S} = \mathbf{U}^T(\mathbf{P}_D \circ [\mathbf{x}\ \mathbf{x}\ \ldots\ \mathbf{x}]), \tag{8}$$

where $[\mathbf{x}\ \mathbf{x}\ \ldots\ \mathbf{x}]$ is a $N \times N$ matrix whose columns are signal vectors $\mathbf{x}$.

For a rectangular window $g(d) = 1$ the LVS can be calculated recursively with respect to the window width D as

$$\mathbf{S}_D = \mathbf{S}_{D-1} + \mathbf{U}^T(\mathbf{A}_{D-1} \circ [\mathbf{x}\ \mathbf{x}\ \ldots\ \mathbf{x}]). \tag{9}$$

Example: The local vertex frequency representation of the signal from Fig. 1 is shown in Fig. 3. The signal consists of parts of three eigenvectors. For the subset $\mathcal{V}_1$ which includes vertices from 1 to 13, eigenvector $k = 20$ is used. For the subset $\mathcal{V}_2$ with vertices 14 to 26, the signal is equal to eigenvector $k = 52$. Remaining vertices form $\mathcal{V}_3$, and the signal on this subset is equal to eigenvector $k = 36$. Amplitudes of the eigenvectors are scaled as well.

The window and modulated window (kernel) used for local vertex-frequency analysis for the vertices $m = 8$ (top) and $m = 31$ (bottom) and spectral indices $k = 1$ (left) and $k = 6$ (right) are shown in Fig. 2.

Comments

1) *Weighted distance:* For a weighted graph, the distance from a vertex n to a vertex m could be defined by using the edge weights instead of the number of edges. For example we can define distance as a sum, or product of the associated edge weights. Then d_{nm} may assume non-integer values.

2) *Interpolation:* Note that for the windowed signal $x(n)h_m(n)$, only $M \leq N$ samples are nonzero. It may be considered a classical zero padded signal. This means that for the reconstruction of this signal, we only need M spectral coefficients $S(n,k)$. The remaining coefficients can be calculated from the system of equations obtained by using the fact that $x(n)h_m(n) = 0$ outside the window support. In the classical signal analysis, it is common to calculate the STFT with M frequency indices for a window whose width is M. The same can be done in graph signal processing. The local vertex spectrum, at vertex m, can be written in a matrix form as

$$\begin{bmatrix} \mathbf{s}_m^{(B)} \\ \mathbf{s}_m^{(I)} \end{bmatrix} = \begin{bmatrix} \mathbf{B}^{(B)} \\ \mathbf{B}^{(I)} \end{bmatrix} \mathbf{x}_{h_m}^{(NZ)},$$

where $\mathbf{x}_{h_m}^{(NZ)}$ is vector with M nonzero elements of windowed signal $\mathbf{x}_{h_m}$, $\mathbf{B}^{(B)}$ is square $(M \times M)$ submatrix of the transformation matrix $\mathbf{U}^T$ with omitted columns that correspond to zero elements in $\mathbf{x}_{h_m}$, while the remaining $N - M$ rows are arranged into matrix $\mathbf{B}^{(I)}$. We have assumed that the local spectrum vector $\mathbf{s}_m$ is composed of two parts $\mathbf{s}_m^{(B)}$ and $\mathbf{s}_m^{(I)}$ in a such way that the matrix $\mathbf{B}^{(B)}$ is invertible. Then

$$\mathbf{s}_m^{(B)} = \mathbf{B}^{(B)}\mathbf{x}_{h_m}^{(NZ)}$$

and

$$\mathbf{s}_m^{(I)} = \mathbf{B}^{(I)}\mathbf{x}_{h_m}^{(NZ)} = \mathbf{B}^{(I)}(\mathbf{B}^{(B)})^{-1}\mathbf{s}_m^{(B)}$$

is the interpolation formula that enables calculation of $\mathbf{s}_m^{(I)}$ from $\mathbf{s}_m^{(B)}$. Note that if the condition number of matrix $\mathbf{B}^{(B)}$ is large, we can rearrange the spectral domain coefficients ordering until we get a satisfactory invertible matrix.

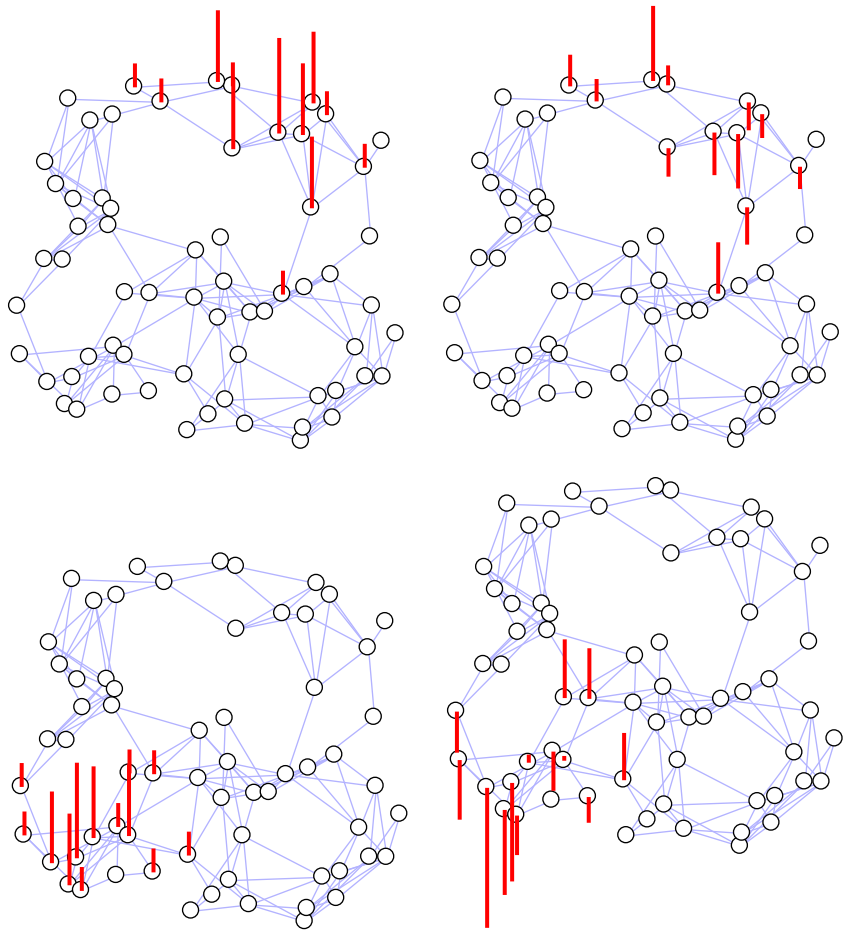

Fig. 2 Kernels with the vertex domain localized windows centered at vertices 8 (top) and 31 (bottom) and spectral indices 1 (left) and 6 (right) (LVS)

3) *Directed graphs:* The vertex neighborhood may be defined as set of vetices that can be reached from the considered vertex by a walk whose length is at most D. Then we may use previous relations for the window calculation. This approach corresponds to one-sided windows in classical signal analysis.

If we want to define two-sided window, then we should also include all vertices from which we can reach the considered vertex by walk whose length is at most D. This means that for a directed graph we should assume that vertices with distance one form the considered vertex m are the vertices from which we can reach vertex m with walk of length 1. In this case $\mathbf{A}_1 = \mathbf{A} + \mathbf{A}^T$ where addition is logical operation (Boolean OR). The matrix $\mathbf{A}_2$ is

$$\mathbf{A}_2 = (\mathbf{A} \odot \mathbf{A} + \mathbf{A}^T \odot \mathbf{A}^T) \circ (\mathbf{1} - \mathbf{I}) \circ (\mathbf{1} - \mathbf{A}_1).$$

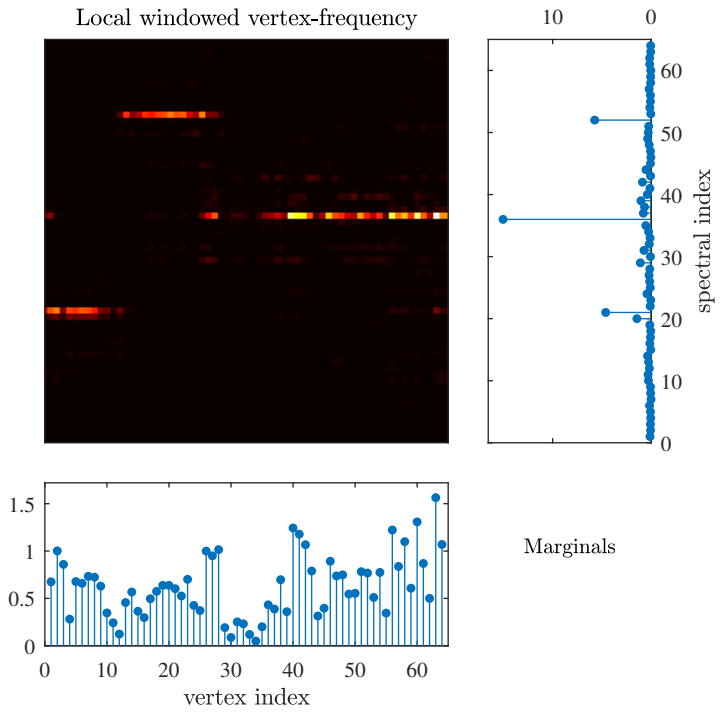

Fig. 3 Local vertex-frequency spectrum calculated using vertex neighborhood windows

This procedure could be continued for walks up to the desired maximal length D.

For a circular directed graph in this way, we will get the classical STFT with symmetric window.

4) *Ordering:* In order to visualize local spectral content, we should order vertices in the corresponding graph. This ordering is not unique, and one possible way is to define the order according to the values of low order eigenvectors of the Laplacian. We can, for example, try to minimize the number of zero crossing in the low order eigenvectors by an appropriate vertex reordering. This can be achieved by reordering vertices such that the elements of $b(n)$ are nondecreasing, where $b(n)$ is defined as

$$b(n) = \sum_{k=2}^{K} (1 + \text{sign}(u_k(n)))2^{-k} \tag{10}$$

where K is the number of considered eigenvectors. Here we consider the sign of the eigenvector coefficients and group coefficients with the same sign. Then the sign of the next eigenvector is used to order coefficients in each group. This ordering is based on the vertex spectral similarity. Note that for $k = 1$ we have $\lambda_1 = 0$ and the corresponding eigenvector is constant.

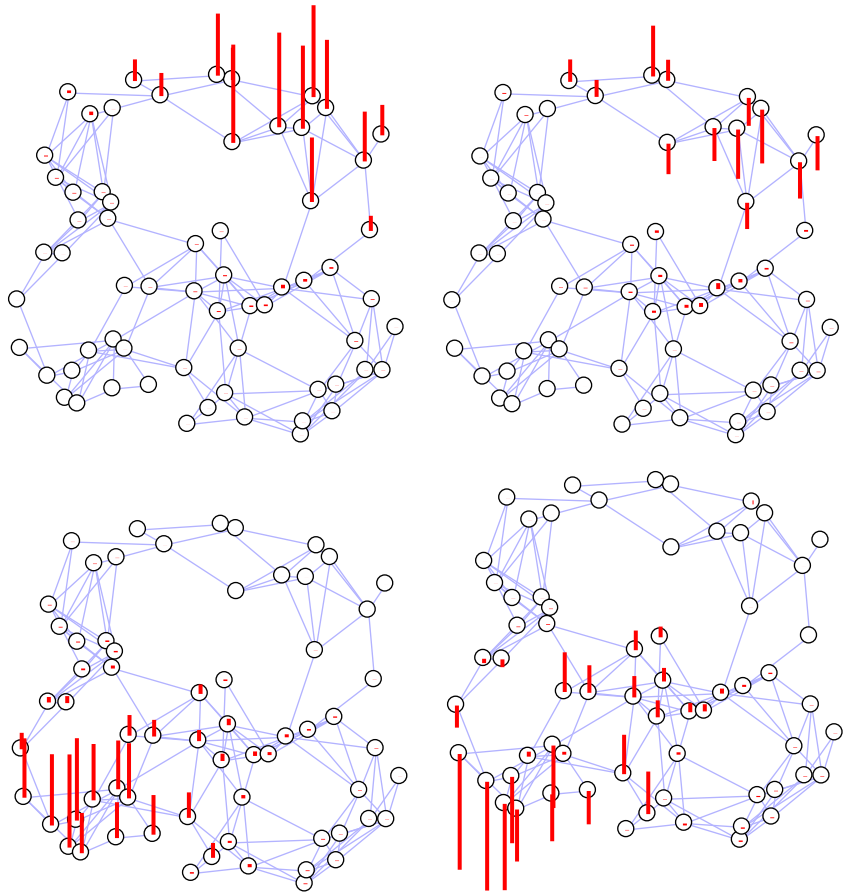

Fig. 4 Kernels with the spectral domain defined windows centered at vertices 8 (top) and 31 (bottom) and spectral indices 1 (left) and 6 (right) (LVS)

2.4 Vertex Localization Windows Defined in the Spectral Domain

Consider two signals $x(n)$ and $y(n)$ on a graph. The GDFT of these signals are given by $X(k)$ and $Y(k)$. The generalized convolution $z(n)$ of signals $x(n)$ and $y(n)$ can be defined in the GDFT domain as

$$Z(k) = X(k)Y(k)$$
$$z(n) = x(n) * y(n).$$

A "shift" on a graph cannot be extended in a direct way from the classical signal processing theory. The generalized convolution is used to define a "shift" of a window on a graph [13]. A "shift" on a graph from vertex m to vertex n would be achieved by using the delta function located at the vertex m, defined as

$$\delta_m(n) = \delta(n - m). \tag{11}$$

The GDFT of this function is given by

$$\Delta(k) = \sum_{n=1}^{N} \delta_m(n) u_k(n) = u_k(m). \tag{12}$$

The window localized at the vertex m is equal to [13]

$$h_m(n) = h(n) * \delta_m(n) = \sum_{k=1}^{N} H(k) u_k(m) u_k(n), \tag{13}$$

where the window basic function $h(n)$ is defined in the spectral domain, for example, as

$$H(k) = C \exp(-\lambda_k \tau), \tag{14}$$

where C is the window amplitude and $\tau > 0$ is a constant that determines the window width. Two windows obtained using this method are presented in Fig. 4 (left).

The window localized at the vertex m satisfies the following properties:

- Symmetry $h_m(n) = h_n(m)$ follows from definition (13).
- Sum of $h_m(n)$ is equal to $H(1)$,

$$\sum_{n=1}^{N} h_m(n) = \sum_{k=1}^{N} H(k) u_k(m) \sum_{n=1}^{N} u_k(n) = \sum_{k=1}^{N} H(k) u_k(m) \delta(k-1)\sqrt{N} = H(1).$$

- Parseval's theorem for $h_m(n)$ is

$$\sum_{n=1}^{N} |h_m(n)|^2 = \sum_{k=1}^{N} |H(k) u_k(m)|^2. \tag{15}$$

The local vertex spectrum can be written as

$$S(m,k) = \sum_{n=1}^{N} x(n) h_m(n)\ u_k(n) = \sum_{n=1}^{N} \sum_{p=1}^{N} x(n) H(p) u_p(m) u_p(n)\ u_k(n). \tag{16}$$

The modulated version of the window (kernel) centered at vertex m and spectral index k is

$$\mathcal{H}_{m,k}(n) = h_m(n) u_k(n) = \Big(\sum_{p=1}^{N} H(p) u_p(m) u_p(n) \Big) u_k(n).$$

It is shown in Fig. 4 for $k = 1$ (left) and $k = 6$ (right) for two vertex locations $m = 8$ (top) and $m = 31$ (bottom). The kernels with the vertex-domain localized windows, at the same vertex index and spectral index locations, are shown in Fig. 2.

Using the kernel notation, the local vertex spectrum, for a given vertex m, is equal to the projection of a signal $x(n)$ onto the kernel $\mathcal{H}_{m,k}(n)$,

$$S(m,k) = \sum_{n=1}^{N} \mathcal{H}_{m,k}(n) x(n).$$

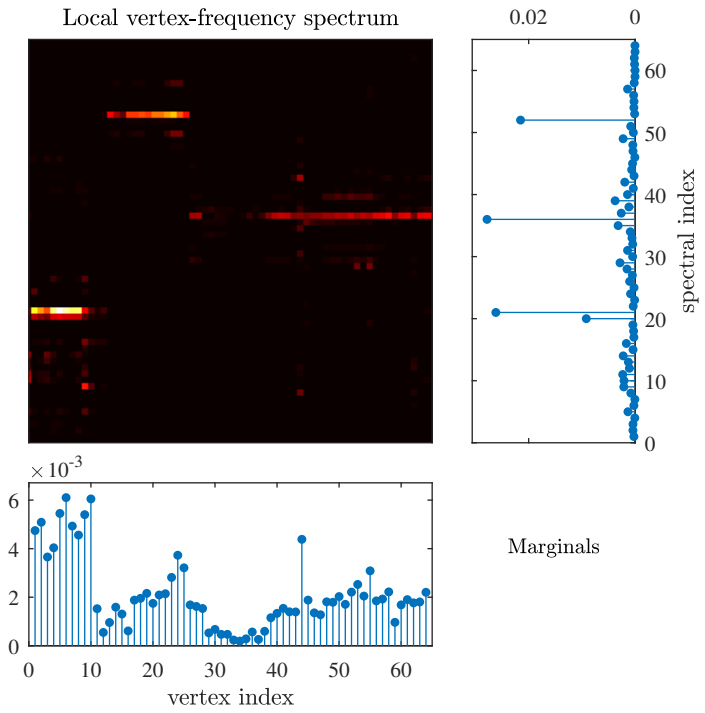

Fig. 5 Local vertex-frequency spectrum calculated using vertex-frequency localized kernels/windows defined in the spectral domain

2.5 Inversion

The inversion relation for both previous window forms can be considered in a unified way. The reconstruction of a signal $x(n)$ from its local spectrum $S(m,k)$ is performed by an inverse GDFT

$$x(n)h_m(n) = \sum_{k=1}^{N} S(m,k)\, u_k(n) \tag{17}$$

followed by a summation over all vertices m

$$x(n) = \frac{1}{\sum_{m=1}^{N} h_m(n)} \sum_{m=1}^{N}\sum_{k=1}^{N} S(m,k) u_k(n). \tag{18}$$

If the windows $h_m(n)$ satisfy the condition

$$\sum_{m=1}^{N} h_m(n) = 1,$$

the reconstruction is vertex independent. In that case

$$x(n) = \sum_{m=1}^{N} \sum_{k=1}^{N} S(m,k) u_k(n) = \sum_{k=1}^{N} X(k) u_k(n), \tag{19}$$

where

$$X(k) = \sum_{m=1}^{N} S(m,k)$$

is the spectral marginal of the local vertex spectrum.

The condition $\sum_{m=1}^{N} h_m(n) = 1$ can be achieved if the elements of matrix $\mathbf{A}_d$, $d = 1, 2, \ldots, D-1$ are normalized, prior to $\mathbf{P}_D$ matrix calculation in such a way that the sum over all columns is equal to 1. Then

$$\sum_{m=1}^{N} h_m(n) = \sum_{m=1}^{N} P_D(n,m) = \sum_{d=1}^{D-1} g(d) = const.$$

For the windows obtained using generalized graph shift, this conditions is always satisfied since $H(1) = const$.

In general, the local vertex spectrum $S(m,k)$ can be calculated over a reduced set of vertices $m \in \mathbb{M} \subset \mathcal{V}$. In this case, summation over m in the reconstruction formula should be performed over vertices $m \in \mathbb{M}$ only, and vertex independent reconstruction is achieved if $\sum_{m \in \mathbb{M}} h_m(n) = 1$.

Inversion of the local vertex spectrum, within the Gabor expansion framework, is obtained from

$$\sum_{m=1}^{N} \sum_{k=1}^{N} S(m,k) \mathcal{H}_{m,k}(n) = \sum_{m=1}^{N} \left(\sum_{k=1}^{N} S(m,k) h_m(n) u_k(n) \right) =$$

$$\sum_{m=1}^{N} \left(\sum_{i=1}^{N} \underset{k \to i}{\text{IGDFT}}\{S(m,k)\} \underset{k \to i}{\text{IGDFT}}\{h_m(n) u_k(n)\} \right) =$$

$$\sum_{m=1}^{N} \sum_{i=1}^{N} x(i) h_m(i) h_m(n) \delta(n-i) = \sum_{m=1}^{N} x(n) h_m^2(n) = x(n) \sum_{m=1}^{N} h_m^2(n),$$

where IGDFT denotes the inverse GDFT transform.

The inversion formula is then

$$x(n) = \frac{1}{\sum_{m=1}^{N} h_m^2(n)} \sum_{m=1}^{N} \sum_{k=1}^{N} S(m,k) h_m(n) u_k(n). \tag{20}$$

This kind of inversion is vertex invariant if

$$\sum_{m=1}^{N} h_m^2(n) = 1. \tag{21}$$

If the local vertex spectrum $S(m,k)$ is calculated over a reduced set of vertices $m \in \mathbb{M} \subset \mathcal{V}$, then vertex independent reconstruction condition is $\sum_{m\in\mathbb{M}} h_m^2(n) = 1$.

Filtering in the vertex-frequency domain can be implemented by using the vertex-frequency support function $B(m,k)$. The filtered local vertex spectrum is

$$S_f(m,k) = S(m,k)B(m,k)$$

and the filtered signal $x_f(n)$ is obtained by inversion of $S_f(m,k)$ using the presented inversion methods.

The support function $B(m,k)$ can be obtained, for example, by thresholding noisy values of the local vertex spectrum $S(m,k)$.

2.6 Uncertainty Principle

In classical signal analysis, the window is used to improve the signal localization in the joint time-frequency domain. However, the uncertainty principle prevents an ideal localization in both time and frequency. The DFT the uncertainty principle states that

$$\|\mathbf{x}\|_0\|\mathbf{X}\|_0 \geq N.$$

It means that the product of the number of signal nonzero values $\|\mathbf{x}\|_0$ and the number of its DFT nonzero values $\|\mathbf{X}\|_0$ is greater or equal than the total number of signal samples N.

If the signal is windowed with a function $\mathbf{h}_m$ whose width is N_{h_m} then $\|\mathbf{x}\mathbf{h}_m\|_0 \leq N_{h_m}$. The uncertainty principle for the classical STFT, defined as $\mathbf{STFT}_m = \mathrm{DFT}\{\mathbf{x}\mathbf{h}_m\}$, is

$$\|\mathbf{STFT}_m\|_0 \geq \frac{N}{N_{h_m}}.$$

This means that the window whose width is, for example, $N_{h_m} = N/4$ can not produce the STFT with less than 4 nonzero samples for considered instant m.

For graph signals, the general form of the uncertainty principle should be used. Consider a graph signal $\mathbf{x}$ and its transform $\mathbf{X}$ in the domain of orthonormal basis functions $u_k(n)$. The uncertainty principle states that [20–23]

$$\|\mathbf{x}\|_0\|\mathbf{X}\|_0 \geq \frac{1}{\max_{k,m}\{|u_k(m)|^2\}}.$$

For the orthonormal DFT, when $u_k(n) = \frac{1}{\sqrt{N}}\exp(j2\pi nk/N)$, the classical DFT uncertainty principle form follows.

In graph signal processing, the basis functions can assume quite different forms than in the DFT case. In a cases, when, for example, there is a vertex loosely connected with other vertices $\max\{|u_k(m)|^2\} \to 1$ and even

$\|\mathbf{x}\|_0\|\mathbf{X}\|_0 \geq 1$ is possible. This means that the graph signal can be well localized in both the vertex and the frequency domain.

For the graph presented in Fig. 1, $\max\{|u_k(m)|^2\} = 0.7513$ meaning that $\|\mathbf{x}\|_0\|\mathbf{X}\|_0 \geq 1.331$ is possible.

2.7 Local Vertex Spectrogram and Energy Condition

The local vertex spectrogram is defined as

$$|S(m,k)|^2 = \Big| \sum_{n=1}^{N} x(n)h_m(n)\ u_k(n)\Big|^2. \tag{22}$$

.

The vertex marginal property is (according to Parseval's theorem)

$$\sum_{k=1}^{N} |S(m,k)|^2 = \sum_{k=1}^{N} S(m,k) \sum_{n=1}^{N} x(n)h_m(n)\ u_k(n) = \sum_{n=1}^{N} |x(n)h_m(n)|^2. \tag{23}$$

This is a vertex smoothed signal power.

For the energy, we get

$$\sum_{m=1}^{N}\sum_{k=1}^{N} |S(m,k)|^2 = \sum_{n=1}^{N} \Big(|x(n)|^2 \sum_{m=1}^{N} |h_m(n)|^2\Big). \tag{24}$$

If $\sum_{m=1}^{N} |h_m(n)|^2 = 1$ for all n then the spectrogram on the graph is energy unbiased

$$\sum_{m=1}^{N}\sum_{k=1}^{N} |S(m,k)|^2 = \sum_{n=1}^{N} |x(n)|^2 = E_x. \tag{25}$$

From the previous relations we can easily prove that the local vertex spectrum is a frame [24–27] since

$$\sum_{m=1}^{N} |h_m(n)|^2 = \sum_{k=1}^{N} |H(k)|^2 |u_k(n)|^2. \tag{26}$$

We can write

$$\frac{1}{N}H^2(1) \leq \sum_{m=1}^{N} |h_m(n)|^2 \leq \max_{m,k}\{|u_k(n)|^2\} \sum_{k=1}^{N} |H(k)|^2 = \mu^2 E_h, \tag{27}$$

where $\mu = \max_{m,k}\{|u_k(n)|\}$. This means that

$$\frac{1}{N}H^2(1)E_x \leq \sum_{m=1}^{N}\sum_{k=1}^{N} |S(m,k)|^2 \leq E_x\mu^2 E_h. \tag{28}$$

The shift $\mathcal{H}_{m,k}(n)$ is an equiangular tight frame (ETF) if $\mu = |u_k(n)| = \frac{1}{\sqrt{N}}$. Then $\sum_{m=1}^{N} |h_m(n)|^2 = \sum_{k=1}^{N} |H(k)|^2 |u_k(n)|^2 = E_h/N$, and equality holds.

2.8 Optimization

The concentration of the local vertex spectrum representation can be measured using the normalized norm-one [28]

$$\mathcal{M} = \frac{1}{F}\sum_{m=1}^{N}\sum_{k=1}^{N}|S(m,k)| = \frac{\|\mathbf{S}\|_1}{\|\mathbf{S}\|_F}, \tag{29}$$

where $F = \|\mathbf{S}\|_F = \sqrt{\sum_{m=1}^{N}\sum_{k=1}^{N}|S(m,k)|^2}$ is the Frobenius norm of matrix $\mathbf{S}$. Any other norm $\|\mathbf{S}\|_p^p$ with $0 \leq p \leq 1$ can be used instead of $\|\mathbf{S}\|_1$. Norms with p close to 0 are noise sensitive. The norm with $p = 1$ is the only convex norm, allowing the gradient based normalization [28].

The concentration measure $\mathcal{M}(\tau) = \|\mathbf{S}\|_1/\|\mathbf{S}\|_F$ for the signal from Fig. 1 and the window given by (14), is shown in Fig. 6, for various τ. The optimal vertex frequency representation is also given in Fig. 6. The optimal τ can be obtained in a few steps in an iterative way $\tau_k = \tau_{k-1} - \alpha(\mathcal{M}(\tau_{k-1}) - \mathcal{M}(\tau_{k-2}))$.

The optimization of parameter τ can be done by using more advanced techniques [21, 22] based on the graph uncertainty principle.

2.9 Spectral Domain Localization

In the classical STFT, it is possible to perform localization using a window in the spectral domain. The dual form of STFT is obtained using inverse DFT. Similarly, for graph signals we can perform localization in the spectral domain and obtain a local-vertex spectrum as an inverse GDFT

$$S(m,k) = \sum_{p=1}^{N} X(p)H(p-k)\,u_k(m), \tag{30}$$

where $H(p-k)$ is the frequency domain window and $X(p)$ is the GDFT of the considered signal.

3 Vertex-Frequency Energy Distributions

Vertex-frequency representations based on the distribution of energy over the vertex and spectral index space is presented in this section. It follows the concept of time-frequency energy distributions in time-frequency signal analysis. An ideal vertex-frequency energy distribution will be discussed first. Special attention will be paid to local smoothness and marginal properties. Two forms of a distribution corresponding to the Rihaczek distribution will be introduced. Finally, a class of reduced interference vertex-frequency distributions, satisfying the marginal properties, will be presented.

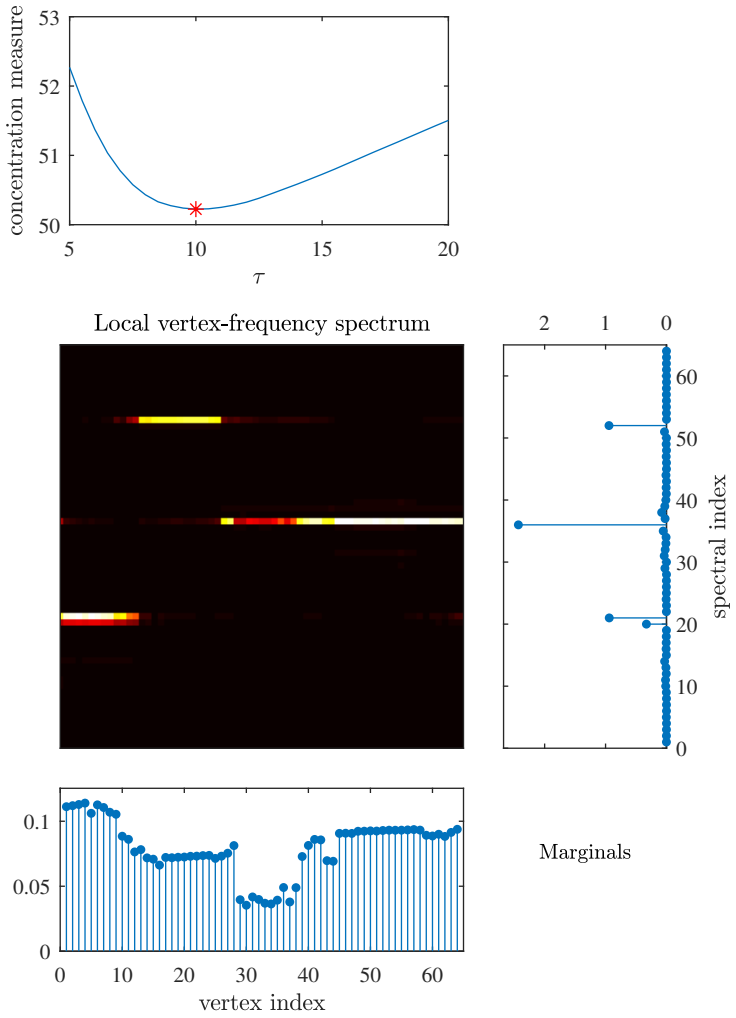

Fig. 6 Measure of the spectrogram concentration for various window parameter τ (top) and the corresponding optimal vertex-frequency representation with marginals (bottom).

3.1 Global Graph Signal Smoothness

The smoothness of a graph signal $\mathbf{x}$ is defined by using its quadratic form

$$E_x = \mathbf{x^T L x}.$$

The smoothness λ_x is equal to the quadratic form normalized by the signal energy

$$\lambda_x = \frac{\mathbf{x}^T\mathbf{L}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}.$$

In the case of classical time domain signals, the Laplacian on a circle graph represents the second order finite difference $y(n) = -x(n-1)+2x(n)-x(n+1)$.

This difference can be written in a matrix form as $\mathbf{y} = \mathbf{L}\mathbf{x}$. It is obvious that the signal $x(n)$ with small changes should have small quadratic form $E_x = \sum_n((x(n)-x(n-1))^2+(x(n)-x(n+1))^2)/2$. This reasoning can be used for the graph signals as well. For these signals $E_x = \sum_{m=1}^{N}\sum_{n=1}^{N} w_{mn}((x(n) - x(m))^2/2$ (see Chapter I).

The eigenvector and eigenvalue relation is $\mathbf{L}\mathbf{u}_k = \lambda\mathbf{u}_k$ or

$$\mathbf{u}_k^T\mathbf{L}\mathbf{u}_k = \lambda_k\mathbf{u}_k^T\mathbf{u}_k = \lambda_k = E_{u_k}, \tag{31}$$

since $\mathbf{u}_k^T\mathbf{u}_k = 1$. The quadratic form of an eigenvector is equal to the corresponding eigenvalue. It can be used as a measure of the signal smoothness. The eigenvectors corresponding to small λ_k belong to the low-pass (slow-varying) part of a graph signal.

Examples of signals with a small, a moderate and a large smoothness λ_x are presented in Fig. 7. In order to make an analogy with the classical time-domain signal processing we presented the time-domain signals in Fig. 7 (left). They can be considered as the signals on a circular graph. The smooth signals, with small smoothness factors, have similar signal values on the neighboring vertices (time instants), while the similarity of neighboring signal values does not hold for fast-varying signals with high smoothness indices.

For a signal of the form

$$x(n) = \sum_{i=1}^{M} x_i(n) = \sum_{i=1}^{M} \alpha_i u_{k_i}(n)$$

the global smoothness is

$$\lambda_x = \frac{\sum_{i=1}^{N}\alpha_i^2\lambda_{k_i}}{\sum_{i=1}^{N}\alpha_i^2}.$$

It is obvious that $\lambda_{\min} \leq \lambda_x \leq \lambda_{\max}$, where $\lambda_{\min} = \min\{\lambda_{k_1}, \lambda_{k_2}, \dots \lambda_{k_M}\}$ and $\lambda_{\max} = \max\{\lambda_{k_1}, \lambda_{k_2}, \dots \lambda_{k_M}\}$.

The graph signal smoothness plays an important role in the graph topology learning. The smoothness is also used in the vertex ordering and the graph clustering.

3.2 Local Graph Signal Smoothness

Consider a single component graph signal

$$x(n) = \alpha u_k(n),$$

where $u_k(n)$ is the kth Laplacian eigenvector. The smoothness of this signal is equal to the corresponding eigenvalue

$$\frac{\mathbf{x}^T\mathbf{L}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \frac{\alpha^2\mathbf{u}_k^T\mathbf{L}\mathbf{u}_k}{\alpha^2\mathbf{u}_k^T\mathbf{u}_k} = \lambda_k.$$

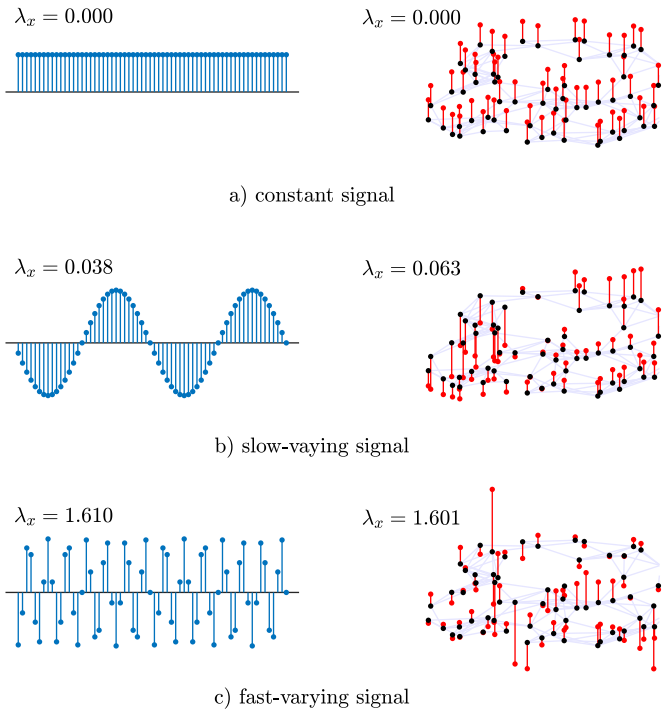

Fig. 7 An example of a constant, a slow-varying, and a fast-varying signal in the time domain (left) and in the graph domain (right). The global signal smoothness λ_x is given for each case.

The smoothness can be related to the frequency in classical signal analysis. For a circular undirected graph the eigenvectors are periodic functions $u_k(n) = \cos(2\pi nk/N + \phi)$, with frequency $\omega = 2\pi k/N$, and the smoothness follows as the eigenvalue from the equation $\mathbf{L}\mathbf{u}_k = \lambda_k \mathbf{u}_k$ as

$$\lambda_k = 4\sin^2(\omega/2) = \mathbf{u}_k^T \mathbf{L}\mathbf{u}_k.$$

This means that the eigenvalue corresponds to the squared frequency. The relation is via $2\sin^2(\omega/2)$ function due to the discretization (discrete-time to continuous-time frequency mapping using the first-order finite difference). In the continuous-time case (for a small ω in the discrete-time domain) we would have $\lambda_k \approx \omega^2$.

In classical signal analysis, for signals whose frequency changes, the instantaneous frequency is defined and used instead of the frequency. Various definitions of the instantaneous frequency are introduced [9–11]. The most straightforward one would be to define the instantaneous frequency as the frequency of a sinusoidal signal which best fits the signal behavior at and in the very close vicinity of the considered instant t. In that case, the method for frequency estimation of the signal in three close points could be used, as described in [29]. The signal $x(t+\tau)$ is approximated with a second order

polynomial around $x(t)$,

$$x(t+\tau) \approx x(t) + x'(t)\tau + x''(t)\tau^2/2.$$

By comparing this signal with a general sinusoidal signal at an instant t and small τ

$$\begin{aligned} A\cos(\omega_t(t+\tau)+\phi) \approx A\cos(\omega_t t+\phi) - A\omega_t \sin(\omega_t t+\phi)\tau \\ - A\omega_t^2 \cos(\omega_t t+\phi)\tau^2/2 \end{aligned}$$

we can conclude that for $x(t) \neq 0$ the sinusoidal signal with frequency

$$\omega_t^2 = \omega^2(t) = \lambda(t) = -\frac{x''(t)}{x(t)}$$

fits the signal around this instant. In this way we can define the instantaneous frequency $\omega^2(t)$ of a signal at an instant t, assuming that the conditions for the second order polynomial approximation holds. For $x(t) = 0$, the instantaneous frequency can be calculated using the higher-order derivatives ratio $x^{(n+2)}(t)/x^{(n)}(t)$ assuming that $x^{(n)}(t) \neq 0$.

The discrete-time form of the instantaneous frequency relation is

$$\lambda(n) = \omega^2(n) = -\frac{x(n-1) - 2x(n) + x(n+1)}{x(n)} = \frac{\mathcal{L}_x(n)}{x(n)}, \tag{32}$$

where $\mathcal{L}_x(n)$ are the elements of vector $\mathbf{Lx}$ (for a circular graph). An example of a time-domain signal that contains a slow-varying component at the beginning, a fast-varying in the middle, and a moderate-varying component at the end is presented in Fig. 8.

The last relation will be used to introduce the local smoothness for a general graph as

$$\lambda(n) = \frac{\mathcal{L}_x(n)}{x(n)}, \tag{33}$$

with the assumption $x(n) \neq 0$.

Some of the local smoothness properties are listed next [30].

1. For a monocomponent signal

   $$x(n) = \alpha u_k(n)$$

   the local smoothness is constant and it is equal to its global smoothness

   $$\lambda(n) = \frac{\alpha \mathcal{L}_{u_k}(n)}{\alpha u_k(n)} = \lambda_k,$$

   since $\mathcal{L}_{u_k}(n) = \lambda_k u_k(n)$ holds for each element of the matrix equation $\mathbf{L}\mathbf{u}_k = \lambda_k \mathbf{u}_k$. In classical signal analysis this means that the instantaneous frequency of a sinusoidal signal (as a basis function) is equal to the component frequency.

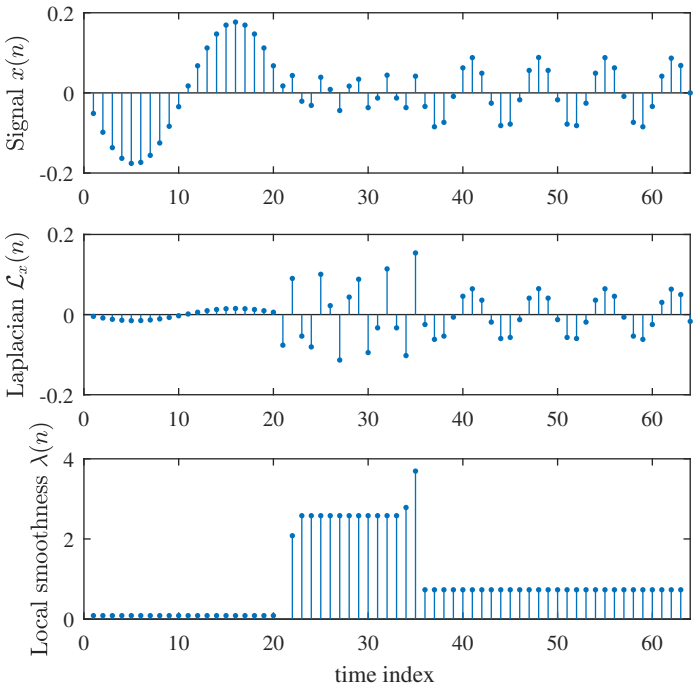

Fig. 8 An example of the signal local smoothness in the time domain (vertex domain on a circular graph).

2. A piecewise monocomponent signal could be defined as

$$x(n) = \alpha_i u_{k_i}(n) \text{ for } n \in \mathcal{V}_i, \quad i = 1, 2, \ldots, M$$

where $\mathcal{V}_i$ are disjunct subsets of vertices and $u_{k_i}(n)$ are eigenvectors. The local smoothness for this signal should be

$$\lambda(n) = \frac{\alpha_i \mathcal{L}_{u_{k_i}}(n)}{\alpha_i u_{k_i}(n)} = \lambda_{k_i} \text{ for } n \in \mathcal{V}_i, \quad i = 1, 2, \ldots, M \tag{34}$$

This is true for all interior vertices, where the vertex n and its neighborhood (used for the Laplacian calculation) belong to the considered subset $\mathcal{V}_i$.
An example of a piecewise monocomponent signal is presented in Fig. 1. Three subsets of vertices $\mathcal{V}_1$, $\mathcal{V}_2$, and $\mathcal{V}_3$ are marked with colors. The component spectral indices are $k_1 = 20$, $k_2 = 52$, and $k_3 = 36$.
For subset $\mathcal{V}_1$, which includes vertices from 1 to 13, the boundary vertices are 1, 2, and 13. The remaining vertices are the interior vertices, where relation (34) holds. For subset $\mathcal{V}_2$ (vertices from 14 to 26), the boundary vertices are 15, 16, 25, and 26. For subset $\mathcal{V}_3$ (vertices from 27 to 64), the boundary vertices are 27, 28, 29, 63 and 64.
The localsmoothness of the considered signal is calculated and presented in Fig. 9. Results obtained for the interior vertices are exact. They are

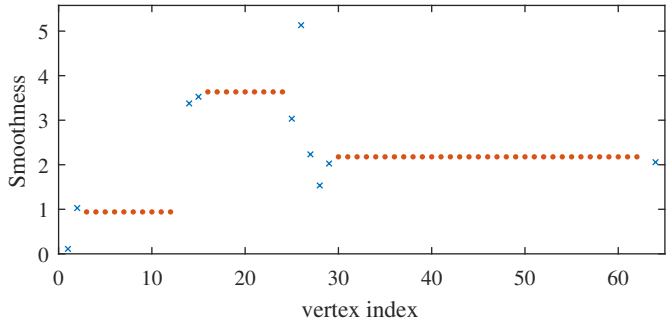

Fig. 9 Local smoothness for the signal given in Fig. 1. The results for the interior vertices are presented with points and the results for the boundary vertices are presented by crosses.

presented with dots. The local smoothness at the boundary vertices is not exact, as expected, since it includes vertices with different signal components. They are presented with the cross marks.

3. An ideal vertex-frequency distribution $I(n,k)$ should behave as

$$I(n,k) \sim |x(n)|^2 \delta(\lambda_k - \lambda(n)),$$

assuming that the local smoothness is equal to an eigenvalue, or that it is rounded to the nearest eigenvalue.
The ideal vertex frequency distribution for the graph and the signal presented in Fig. 1 is shown in Fig. 10.
We can conclude, that a distribution, behaving as an ideal vertex-frequency distribution, can be used as an estimator of the local smoothness as

$$\hat{\lambda}(n) = \arg\max_k \{I(n,k)\}.$$

This estimator is common and widely used in classic time-frequency analysis [9–11].

4. For an M component graph signal $x(n) = \sum_{i=1}^{M} x_i(n) = \sum_{i=1}^{M} \alpha_i u_{k_i}(n)$, the local smoothness is

$$\lambda(n) = \frac{\sum_{i=1}^{M} \alpha_i \mathcal{L}_{u_{k_i}}(n)}{\sum_{i=1}^{M} \alpha_i u_{k_i}(n)} = \frac{\sum_{i=1}^{M} \alpha_i \lambda_{k_i} u_{k_i}(n)}{\sum_{i=1}^{M} \alpha_i u_{k_i}(n)}.$$

The ideal vertex-frequency representation should not be based on the local smoothness $\lambda(n)$ of the complete multicomponent signal, but on the smoothness of each individual signal component $x_i(n)$ denoted by $\lambda_i(n)$. Its form is

$$I(n,k) \sim \sum_{i=1}^{M} |x_i(n)|^2 \delta(\lambda_k - \lambda_i(n)).$$

The concept of ideal vertex-frequency distribution can be extended to piecewise multicomponent signals.

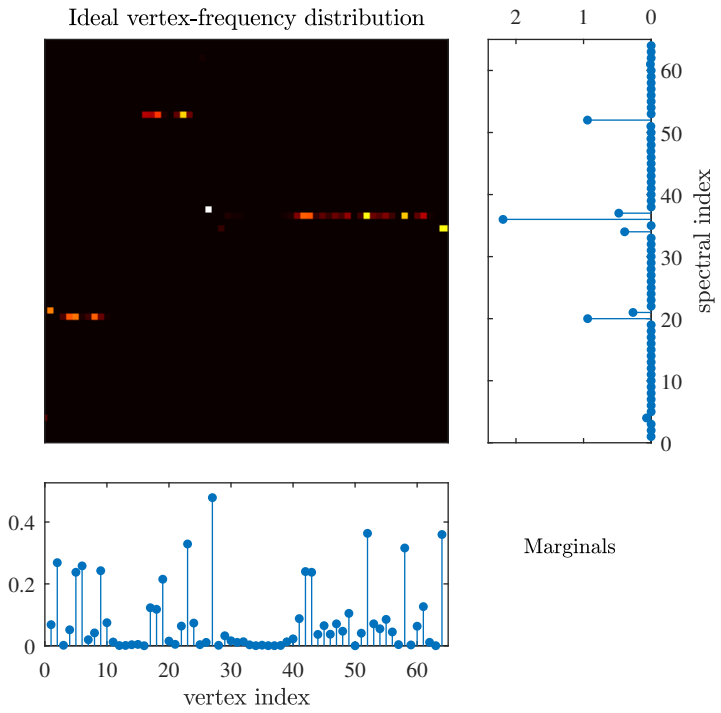

Fig. 10 Ideal vertex-frequency distribution.

5. The local smoothness property holds for a general vertex-frequency distribution $G(n,k)$ if

$$\frac{\sum_{k=1}^{N}\lambda_k G(n,k)}{\sum_{k=1}^{N} G(n,k)} = \lambda(n). \tag{35}$$

Example: The ideal time-frequency distribution $I(n,k) = |x(n)|^2\delta(\lambda_k - \lambda(n))$ satisfies the local smoothness property if $\lambda(n) \in \{\lambda_1, \lambda_2, \ldots, \lambda_N\}$ for all n.

6. The local smoothness bandwidth for a vertex-frequency distribution $G(n,k)$ that satisfies Property 5 is defined by

$$\begin{aligned}\sigma_\lambda^2(n) &= \frac{\sum_{k=1}^{N}(\lambda_k - \lambda(n))^2 G(n,k)}{\sum_{k=1}^{N} G(n,k)} \\ &= \frac{\sum_{k=1}^{N}\lambda_k^2 G(n,k)}{\sum_{k=1}^{N} G(n,k)} - \lambda^2(n).\end{aligned} \tag{36}$$

3.3 Vertex-Frequency Power Distribution

We will present two forms of vertex-frequency distributions. For the first form, we will follow the electric circuit reasoning, while for the second form, the signal processing definition of energy will be used.

3.4 Vertex-Frequency Power in Electric Circuit

Consider a resistive electric circuit and the corresponding graph, where the edge weights w_{nm} are equal to the corresponding conductances $1/R_{nm}$. The potential at the vertex n is denoted by $x(n)$. The power in all edges connected to the vertex n is equal to the sum of all $(x(n)-x(j))^2/R_{nj}$ or $w_{nj}(x(n)-x(j))^2$. Its value is

$$p(n)=\sum_{j=1}^{N} w_{nj}(x(n)-x(j))^2.$$

The power within the whole network is

$$P=\sum_{n=1}^{N}\sum_{j=1}^{N}\frac{1}{2}w_{nj}(x(n)-x(j))^2.$$

The factor $\frac{1}{2}$ is the result of the fact that all edges are taken twice in the summation over all vertices in the circuit. This relation can also be obtained by introducing the external current generators $i_G(n)$ at each vertex. These generators are needed to obtain the actual potentials $x(n)$. The vector of all external currents is denoted by $\mathbf{i}_G$. According to Kirchoff's first law $\mathbf{i}_G=\mathbf{Lx}$. The total power in a circuit is then calculated as

$$P=\mathbf{x}\mathbf{i}_G=\mathbf{x}(\mathbf{Lx})=\sum_{n=1}^{N}x(n)\sum_{j=1}^{N}w_{nj}(x(n)-x(j)).$$

It can be written as

$$P=\sum_{n=1}^{N}\sum_{k=1}^{N}\sum_{j=1}^{N}\frac{1}{2}w_{nj}(x(n)-x(j))X(k)(u_k(n)-u_k(j)).$$

The total power is obtained as a sum of the terms

$$P(n,k)=\sum_{j=1}^{N}\frac{1}{2}w_{nj}(x(n)-x(j))X(k)(u_k(n)-u_k(j))$$

over the vertex and frequency indices as

$$P=\sum_{n=1}^{N}\sum_{k=1}^{N}P(n,k).$$

Therefore, the value of $P(n,k)$ can be considered as a vertex-frequency power distribution of signal $x(n)$ over this graph, The marginal properties of this distribution are:

$$\sum_{n=1}^{N} P(n,k) = \lambda_k |X(k)|^2 = X_D^2(k) \text{ and}$$
$$\sum_{k=1}^{N} P(n,k) = \sum_{j=1}^{N} \frac{1}{2} w_{nj}(x(n)-x(j))^2 = x_D^2(n), \tag{37}$$

where $x_D^2(n) = \mathcal{L}_x(n)x(n)$.

We have obtained that the spectral power is of the form $\lambda_k|X(k)|^2$. For $k = 1$, this power is zero-valued, since $\lambda_1 = 0$ and the corresponding eigenvector $u_1(n)$ is constant. A constant potential does not produce any power in the network, since the voltage between each pair of vertices is 0. This kind of power, proportional to the frequency (squared), is present in the Teager energy operator.

We can define inverse Laplacian of a signal using the transform $X^2(k)/\lambda_k$. Since the Laplacian of a signal, with the transform $\lambda_k X(k)$, is a kind of second order derivation on a graph, its inverse can be considered as a kind of (double) integration on a graph.

Example: Consider the graph signal from Fig. 1. Its vertex-frequency power distribution is shown in Fig. 11. The marginal values of this distribution (37) are exact.

3.5 Signal Energy Vertex-Frequency Distributions

Energy in signal processing is commonly defined as

$$E = \sum_{n=1}^{N} x^2(n) = \sum_{n=1}^{N} x(n) \sum_{k=1}^{N} X(k)u_k(n).$$

It can be written as

$$E = \sum_{n=1}^{N}\sum_{k=1}^{N} x(n)X(k)u_k(n) = \sum_{n=1}^{N}\sum_{k=1}^{N} E(n,k),$$

where the vertex-frequency energy distribution is defined by [31]

$$E(n,k) = x(n)X(k)u_k(n) = \sum_{m=1}^{N} x(n)x(m)u_k(m)u_k(n). \tag{38}$$

This corresponds to the Rihaczek distribution in time-frequency analysis.

The marginal properties of this distribution are

$$\sum_{n=1}^{N} E(n,k) = |X(k)|^2 \quad \text{and} \quad \sum_{k=1}^{N} E(n,k) = x^2(n).$$

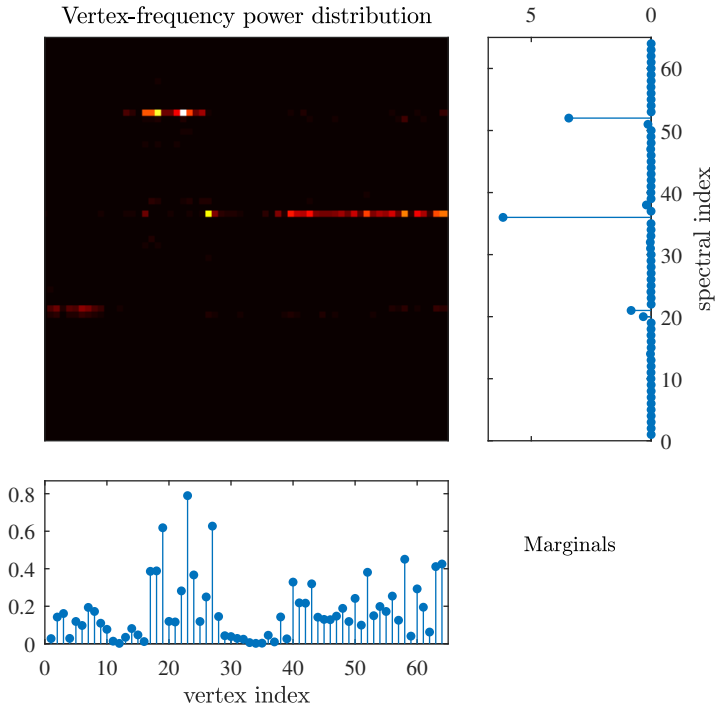

Fig. 11 Vertex-frequency power distribution.

They correspond to the signal power and squared spectra of the graph signal $x(n)$.

The vertex-frequency distribution defined by $E(n,k) = x(n)X(k)u_k(n)$ satisfies the local smoothness property (35). For this distribution

$$\frac{\sum_{k=1}^{N}\lambda_k E(n,k)}{\sum_{k=1}^{N}E(n,k)} = \frac{\sum_{k=1}^{N}\lambda_k x(n)X(k)u_k(n)}{\sum_{k=1}^{N}x(n)X(k)u_k(n)} = \frac{x(n)\mathcal{L}_x(n)}{x^2(n)} = \frac{\mathcal{L}_x(n)}{x(n)} = \lambda(n),$$

since $\sum_{k=1}^{N}\lambda_k X(k)u_k(n)$ are the elements of the inverse GDFT of $\mathbf{\Lambda X}$. This inverse transform is equal to

$$\mathbf{U\Lambda X} = \mathbf{U\Lambda}(\mathbf{U}^T\mathbf{U})\mathbf{X} = (\mathbf{U\Lambda U}^T)(\mathbf{U})\mathbf{X}) = \mathbf{Lx}$$

with elements $\mathcal{L}_x(n)$. Therefore, $\sum_{k=1}^{N}\lambda_k X(k)u_k(n) = \mathcal{L}_x(n)$.

For this vertex-frequency energy distribution, the local smoothness bandwidth (36) may easily be written in terms of the elements of $\mathbf{L}^2\mathbf{x}$, $\mathbf{Lx}$, and $\mathbf{x}$, since $\sum_{k=1}^{N}\lambda_k^2 X(k)u_k(n) = \mathcal{L}_{\mathcal{L}_x}(n)$.

Example: The distribution $E(n,k)$ of the graph signal from Fig. 1, along with the marginal properties, is shown in Fig. 12. The marginal properties are satisfied up to the computer's precision. The localization of energy is better than in the cases obtained with the localized windows in Figs. 5 and 11. This

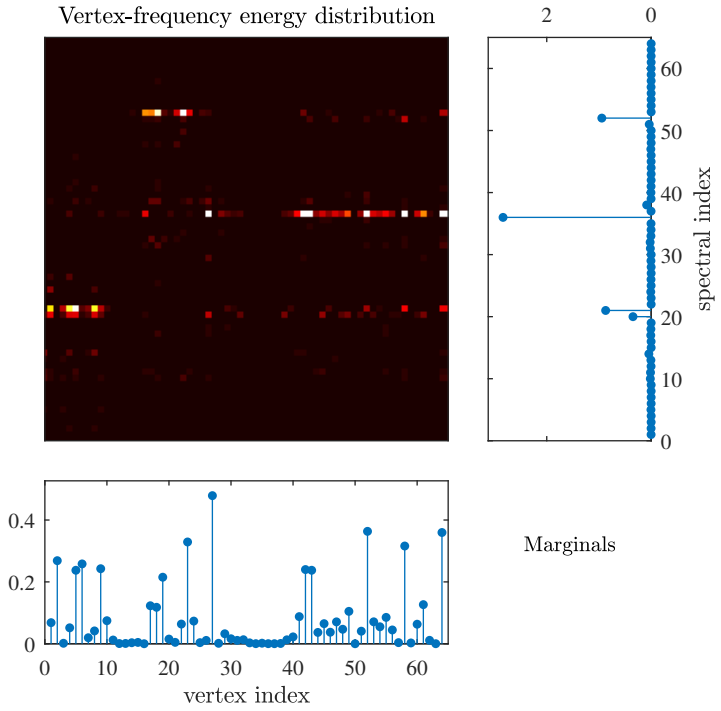

Fig. 12 Vertex-frequency energy distribution.

distribution does not use a localization window. Although the first component of the signal analyzed in Fig. 12 exists in vertices $n \in \mathcal{V}_1 = \{1, 2, \ldots, 13\}$ only, the distribution $E(n, k)$ is nonzero for higher vertex indices n at $k_1 = 20$. This is the cross-term effect, well-known in classical time-frequency analysis. The same effect can be seen at higher vertex indices n at $k_1 = 52$ as well.

4 Reduced Interference Vertex-Frequency Energy Distributions

The general class of time-frequency energy distributions is extended to graph signals in this section with the aim to reduce the cross-term interferences, while preserving the marginal properties [32]. After a review of the classical Cohen class of distribution, conditions for the vertex-frequency marginal properties are derived. Few examples of the vertex-frequency energy distributions are given.

4.1 Review of the classical Cohen class of distributions

Although it is known that any distribution can be used as the basis for the Cohen class of distribution, the Wigner distribution is commonly used [9–11].

Having in mind that the Wigner distribution is not suitable for the graph framework extension, here we will use the Rihaczek distribution as the basis. Since this kind of the Cohen class of distributions is not presented in common literature on time-frequency analysis, a short review of the Cohen class of distributions is presented. The Rihaczek distribution is [9–11]

$$R(t,\omega) = x(t)X^*(\omega)\exp(-j\omega t).$$

Its ambiguity domain form (a two-dimensional Fourier transform of $R(t,\omega)$ over t and ω) is

$$A(\theta,\tau) = \frac{1}{2\pi}\int_u X(u)X^*(u-\theta)\exp(j(u-\theta)\tau)du.$$

The Cohen class of distributions, with the Rihaczek distribution as the basic distribution, is defined by

$$C(t,\omega) = \frac{1}{2\pi}\int_\theta\int_\tau A(\theta,\tau)c(\theta,\tau)\exp(-j\omega\tau)\exp(j\theta t)d\tau d\theta,$$

where $c(\theta,\tau)$ is the kernel function. Using the defined ambiguity domain form of the Rihaczek distribution $A(\theta,\tau)$ we get

$$C(t,\omega) = \frac{1}{4\pi^2}\int_u\int_v X(u)X^*(v)e^{jut}e^{-jvt}\int_\tau c(u-v,\tau)e^{-j\tau\omega}e^{j\tau u}d\tau dudv. \tag{39}$$

The frequency-frequency domain form of the Cohen class of distributions, with the Rihaczek distribution as the basis, is

$$C(t,\omega) = \int_u\int_v X(u)X^*(v)e^{jut}e^{-jvt}\phi(u-v,\omega-u)\frac{dudv}{4\pi^2},$$

where

$$\phi(u-v,\omega-u) = \int_\tau c(u-v,\tau)e^{-j\tau\omega}e^{j\tau u}d\tau.$$

The marginal properties are met if the kernel $c(\theta,\tau)$ satisfies the conditions

$$c(\theta,0) = 1 \text{ and } c(0,\tau) = 1.$$

4.2 Reduced Interference Distributions on Graphs

We will first consider the frequency-frequency domain of the general energy distributions satisfying the marginal properties. The frequency domain definition of the presented energy distribution (38) is

$$E(n,k) = x(n)X^*(k)u_k^*(n) = \sum_{p=1}^{N} X(p)X^*(k)u_p(n)u_k^*(n).$$

Although the basis functions are commonly real-valued, here we used complex-valued notation for possible generalization to the adjacency matrices and directed graphs.

Therefore, the general graph distribution form is

$$G(n,k)=\sum_{p=1}^{N}\sum_{q=1}^{N}X(p)X^{*}(q)u_p(n)u_q^{*}(n)\phi(p,k,q). \tag{40}$$

For $\phi(p,k,q)=\delta(q-k)$ the graph Rihaczek distribution (38) follows. The unbiased energy condition

$$\sum_{k=1}^{N}\sum_{n=1}^{N}G(n,k)=E_x$$

is satisfied if

$$\sum_{k=1}^{N}\phi(p,k,p)=1.$$

The distribution $G(n,k)$ may satisfy the vertex and frequency marginal properties:

– The vertex marginal property is satisfied if

$$\sum_{k=1}^{N}\phi(p,k,q)=1$$

since

$$\sum_{k=1}^{N}G(n,k)=\sum_{p=1}^{N}\sum_{q=1}^{N}X(p)X^{*}(q)u_p(n)u_q^{*}(n)=|x(n)|^2.$$

The same condition is required for the vertex moment property

$$\sum_{n=1}^{N}\sum_{k=1}^{N}n^m G(n,k)=\sum_{n=1}^{N}n^m|x(n)|^2.$$

– The frequency marginal property is satisfied if

$$\phi(p,k,p)=\delta(p-k).$$

Then the sum over vertex index produces

$$\sum_{n=1}^{N}G(n,k)=\sum_{p=1}^{N}|X(p)|^2\phi(p,k,p)=|X(k)|^2,$$

since

$$\sum_{n=1}^{N}u_p(n)u_q^{*}(n)=\delta(p-q),$$

that is, the eigenvectors are orthonormal. If the frequency marginal property holds, then the frequency moment property holds as well,

$$\sum_{n=1}^{N}\sum_{k=1}^{N} k^m G(n,k) = \sum_{k=1}^{N} k^m |X(k)|^2.$$

The reduced interference vertex-frequency distribution $G(n,k)$ satisfies the local smoothness property (35) if

$$\frac{\sum_{k=1}^{N} \lambda_k G(n,k)}{\sum_{k=1}^{N} G(n,k)} = \frac{\mathcal{L}_x(n)}{x(n)} = \lambda(n). \tag{41}$$

This means that

$$\frac{\sum_{k=1}^{N}\sum_{p=1}^{N}\sum_{q=1}^{N} X(p)X^*(q)u_p(n)u_q^*(n)\lambda_k\phi(p,k,q)}{\sum_{k=1}^{N}\sum_{p=1}^{N}\sum_{q=1}^{N} X(p)X^*(q)u_p(n)u_q^*(n)\phi(p,k,q)} = \lambda(n)$$

if

$$\sum_{k=1}^{N}\phi(p,k,q) = 1 \qquad \text{and} \qquad \sum_{k=1}^{N}\lambda_k\phi(p,k,q) = \lambda_p.$$

4.3 Reduced Interference Distribution Kernels

A few examples of the reduced interference kernels that satisfy marginal properties, will be presented next.

Choi-Williams (exponential) kernel: The classic form of this kernel is

$$c(\theta,\tau) = \exp(-\theta^2\tau^2/(2\sigma^2)).$$

The frequency-frequency form of this kernel is

$$\phi(\theta,\omega) = \mathrm{FT}_\tau\{c(\theta,\tau)\} = \exp(-\omega^2\sigma^2/(2\theta^2))|\sigma/\theta|\sqrt{2\pi}.$$

Its shifted version would be

$$\phi(u-v,\omega-u) = \frac{\sigma\sqrt{2\pi}}{|v-u|}\exp\Big(-\sigma^2\frac{(\omega-u)^2}{2(v-u)^2}\Big).$$

A straightforward extension to the graph signal processing would be to use the relation $\lambda \sim \omega^2$, with appropriate exponential kernel normalization. We have implemented this form and concluded that it produces results similar to

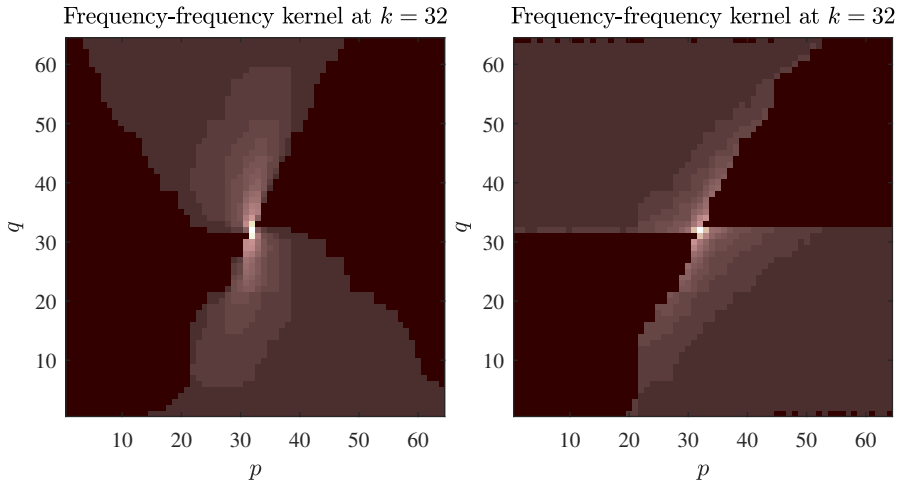

Fig. 13 Frequency-frequency domain exponential (left) and sinc (right) kernels at $k = N/2 = 32$.

the simplified form that satisfies the marginal properties and decays in the frequency-frequency domain. The form of this kernel is

$$\phi(p,k,q) = \frac{1}{s(q,p)} \exp\Big(-\alpha \frac{|\lambda_p - \lambda_k|}{|\lambda_p - \lambda_q|}\Big),$$

where

$$s(q,p) = \sum_{k=1}^{N} \exp\Big(-\alpha \frac{|\lambda_p - \lambda_k|}{|\lambda_p - \lambda_q|}\Big)$$

for $q \neq p$ and $\phi(p,k,p) = \delta(k-p)$. It satisfies both marginal properties.

The vertex-frequency distribution with the exponential kernel (Fig. 13 (left)) is presented in Fig. 14. This kind of distribution presents correctly the signal components, preserving the marginal properties and reducing the cross-term effects as compared to Fig. 12.

Sinc kernel: The simplest reduced interference kernel in the frequency-frequency domain, that would satisfy the marginal properties, is the sinc kernel. Its form is

$$\phi(p,k,q) = \begin{cases} \frac{1}{1+2|p-q|} & \text{for } |k-p| \leq |p-q| \\ 0 & \text{otherwise,} \end{cases}$$

This kernel, with appropriate normalization, is shown in Fig. 13 (right), for $k = 32$. A vertex-frequency representation with this kernel would be similar to the one shown in Fig. 14. Additional examples can be found in [32].

Separable kernels

If the kernel is separable, such that

$$\phi(p,k,q) = g(k-p)g(k-q),$$

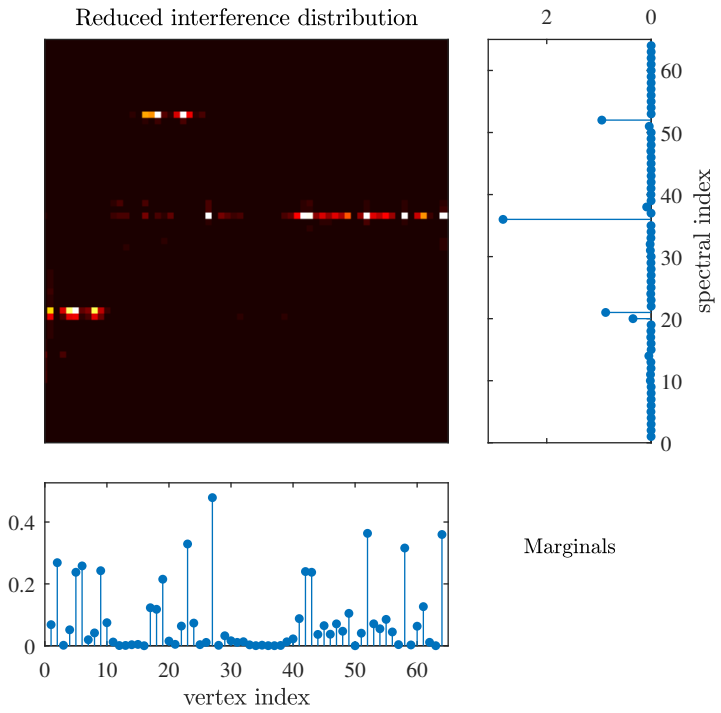

Fig. 14 Vertex-frequency reduced interference distribution using the kernel from Fig. 13(left) with its marginal values equal to $|x(n)|^2$ and $|X(k)|^2$, respectively.

then we can write $G(n,k) = |\sum_{p=1}^{N} X(p)g(k-p)u_p(n)|^2$. This is a frequency domain definition of the graph spectrogram. The relation between the vertex domain spectrogram (6) and the frequency-frequency domain distribution is complex.

The separable kernels cannot satisfy the marginal properties, since $\delta(k-p) = \phi(p,k,p) = g^2(k-p)$ means $g(k-p) = \delta(k-p)$. These kernels do not satisfy $\sum_{k=1}^{N} \phi(p,k,q) = 1$ for all p and q.

Vertex-vertex shift domain distribution

The general vertex-frequency distribution can be written for the vertex-vertex shift domain as a dual form to (40)

$$G(n,k) = \sum_{m=1}^{N}\sum_{l=1}^{N} x(m)x^*(l)u_k(m)u_k^*(l)\varphi(m,n,l), \tag{42}$$

where $\varphi(m,n,l)$ is the kernel in this domain (the same mathematical form as the frequency-frequency domain kernel). The frequency marginal is satisfied if

$$\sum_{n=1}^{N} \varphi(m,n,l) = 1$$

holds. The vertex marginal is met if

$$\varphi(m,n,m)=\delta(m-n).$$

The relation of this distribution with the vertex domain spectrogram (6) is simple using

$$\varphi(m,n,l)=h_n(m)h_n^*(l)$$
$$=\sum_{p=1}^{N}\sum_{q=1}^{N}H(p)H^*(q)u_p(m)u_p(n)u_q^*(l)u_q^*(n).$$

This kernel is defined by the frequency domain window form $H(p)$. It cannot satisfy both marginal properties. The unbiased energy condition

$$\sum_{n=1}^{N}\varphi(m,n,m)=1$$

reduces to (21).

Classical time-frequency analysis

The approach presented in this chapter can be extended to the directed graphs and adjacency matrices as well. The classical Fourier and time-frequency analysis follow from a directed ring graph. The adjacency matrix decomposition produces complex-valued eigenvectors of form $u_k(n)=\exp(j2\pi nk/N)/\sqrt{N}$.

4.4 Local Smoothness Estimation

For the considered signal, presented in Fig. 1, we calculate local signal smoothness by definition (33) and as the maximum positions

$$\hat{\lambda}(n)=\arg\max_k\{G(n,k)\}$$

of the various vertex-frequency representations. The results are presented in Fig. 15 and in Table 1. The local smoothness is calculated only at the vertices where signal has significant values (with instantaneous power $|x(n)|^2$ higher than 0.03 of the maximal instantaneous power $\max_n|x(n)|^2$). The exact local smoothness is presented with a line and the estimated local smoothness at the considered vertices is represented by dots. The number of outliers (vertices where estimated smoothness is not equal to the exact one) is calculated. The mean squared error of the estimations is also presented in each considered case.

Next we considered the same signal and graph example with a white Gaussian noise added to the signal samples. The signal-to-noise ratio (SNR) is 3.6dB. The results are given in Table 2. We can see that the local smoothness estimation based on the signal Laplacian is very sensitive to noise, while

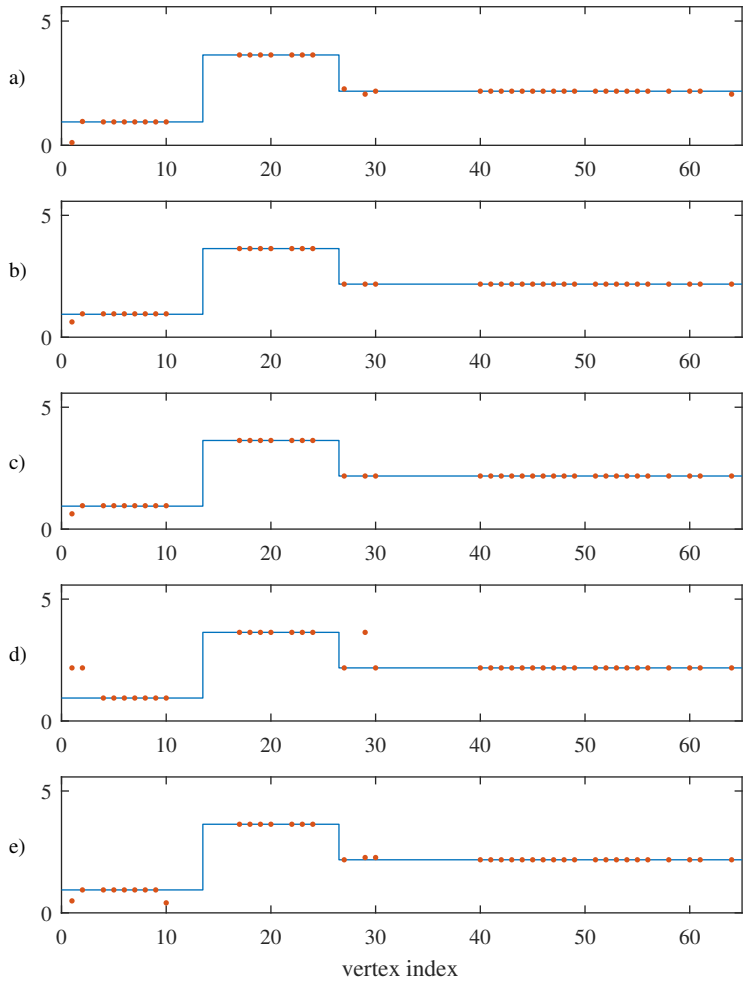

Fig. 15 Local signal smoothness for the graph signal presented in Fig. 1 estimated by using: a) the Laplacian of the graph signal (33), b) the vertex-frequency energy distribution (Fig. 12), c) the reduced interference vertex frequency distribution (Fig. 14), d) the graph spectrogram with a vertex domain window (Fig. 3), and e) the graph spectrogram with a spectral domain window (Fig. 5). The smoothness, presented with dots, is calculated only at the vertices where $|x(n)|^2 > 0.03 \max_n |x(n)|^2$. Solid line is the exact local smoothness in the considered example.

the vertex-frequency based estimations are robust with a slightly increased number of outliers in the noisy case.

Concentration measures obtained as the ℓ_1-norm of the Rihaczek and the reduced interference distributions [28] are 19.63 and 13.26, respectively. They confirm the fact that the reduced interference distribution has improved the concentration in the vertex-frequency domain.

Table 1 The number of outliers and mean squared error of the local smoothness for a noisy free signal.

Calculation method	number of outliers	MSE
Laplacian of the signal	4	0.019
Rihaczek's distribution	1	0.003
Reduced interference distribution	1	0.003
LVS with vertex domain window	3	0.133
LVS with spectral domain window	4	0.013

Table 2 The number of outliers and the mean squared error of the local smoothness for a noisy signal with SNR=3.6dB.

Calculation method	number of outliers	MSE
Laplacian of the signal	36	0.874
Rihaczek's distribution	5	0.119
Reduced interference distribution	3	0.019
LVS with vertex domain window	4	0.087
LVS with spectral domain window	8	0.120

5 Conclusion

Vertex-frequency representations of graph signals have been presented. In the first part of the chapter, the linear signal forms, along with the corresponding window forms and spectrograms, are analyzed. In the second part of the chapter, the local smoothness factor is introduced. The energy vertex-frequency distributions are presented. These distributions do not require localization windows. Energy distributions are extended to the general reduced interference distribution class. This class of graph signal distributions reduces the cross-terms and satisfies the graph signal marginal properties. The theory is illustrated through examples.

References

1. A. Sandryhaila and J. M. F. Moura, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 80–90, Sept 2014.
2. S. Chen, R. Varma, A. Sandryhaila and J. Kovačević, "Discrete Signal Processing on Graphs: Sampling Theory," *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6510-6523, Dec.15, 2015.
3. A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE. Trans. Signal Processing*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.
4. A. Sandryhaila and J. M. F. Moura,"Discrete signal processing on graphs: Frequency analysis," *IEEE. Trans. Signal Processing*, vol. 62, no. 12, pp. 3042–3054, Jun. 2014.
5. A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Stationary Graph Processes and Spectral Estimation," *IEEE Trans. on Signal Processing*, Vol. 65, DOI: 10.1109/TSP.2017.2739099, 2017.
6. D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis

to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, May 2013.
7. I. Jestrović, J. L. Coyle, and E. Sejdić, "Differences in brain networks during consecutive swallows detected using an optimized vertex–frequency algorithm," *Neuroscience*, vol. 344, pp. 113–123, 2017.
8. I. Jestrović, J. L. Coyle, and E. Sejdić, "A fast algorithm for vertex-frequency representations of signals on graphs," *Signal Processing*, vol. 131, pp. 483 – 491, 2017.
9. L. Stanković, M. Daković, and T. Thayaparan, *Time-Frequency Signal Analysis with Applications*, Artech House, Boston, March 2013.
10. L. Cohen, "*Time-frequency Analysis*" Prentice Hall PTR, 1995.
11. B. Boashash, ed, *Time-Frequency Signal Analysis and Processing, A Comprehensive Reference*, Academic Press, 2015.
12. D. I. Shuman, B. Ricaud, P. Vandergheynst, "A windowed graph Fourier transform," *SSP Workshop*, Aug. 2012, pp. 133-136.
13. D. I. Shuman, B. Ricaud, and P. Vandergheynst, "Vertex-frequency analysis on graphs," *Applied and Computational Harmonic Analysis*, vol. 40, no. 2, pp. 260–291, 2016.
14. X. W. Zheng, Y. Y. Tang, J. T. Zhou, H. L. Yuan, Y.L. Wang, L. N. Yang, J.J. Pan, "Multi-windowed graph Fourier frames," *Machine Learning and Cybernetics (ICMLC), 2016 Int. Conf.*, Vol. 2, July, 2016, pp. 1042-1048.
15. M. Tepper, S. Guillermo, "A short-graph fourier transform via personalized pagerank vectors," *Proc. IEEE ICASSP*, 2016.
16. L. Stanković, M. Daković, and E. Sejdić, "Vertex-Frequency Analysis: A Way to Localize Graph Spectral Components," *IEEE Signal Processing Mag.*, Vol.34, pp. 176–182, July 2017.
17. S. Sardellitti, S. Barbarossa, and P. Di Lorenzo, "On the graph Fourier transform for directed graphs," *IEEE Journal of Selected Topics in Signal Processing*, Vol. 11, No. 6, pp. 796–811, 2017.
18. J. A. Deri, and J. M. Moura, "Spectral projector-based graph Fourier transforms," *IEEE Journal of Selected Topics in Signal Processing*, Vol. 11, No. 6, pp. 785–795, 2017.
19. R. Shafipour, A. Khodabakhsh, G. Mateos, and E. Nikolova, "A digraph Fourier transform with spread frequency components," 5th IEEE Global Conf. on Signal and Information Proc. November 14–16, 2017 Montreal, Canada, arXiv preprint arXiv:1705.10821
20. M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2558-2567, Sept. 2002. doi: 10.1109/TIT.2002.801410
21. M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Trans. on Signal Processing*, Vol. 64, No. 18, pp. 4845–4860, 2016.
22. A. Agaskar, and Y. M. Lu, "A Spectral Graph Uncertainty Principle," *IEEE Trans. Information Theory*, Vol. 59, No. 7, pp. 4338–4356, 2013.
23. N. Perraudin, B. Ricaud, D. Shuman, and P. Vandergheynst, "Global and local uncertainty principles for signals on graphs," *APSIPA Transactions on Signal and Information Processing*, 7, E3. doi:10.1017/ATSIP.2018.2, 2018.
24. H. Behjat, U. Richter, D. Van De Ville, L. Sornmo, "Signal-adapted tight frames on graphs," *IEEE Trans. Signal Process.*, Vol.64(22), pp. 6017–6029, 2016.
25. D. Hammond, P. Vandergheynst, R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.* Vol.30(2), pp.129–150, 2011.
26. A. Sakiyama, Y. Tanaka, "Oversampled graph Laplacian matrix for graph filter banks," *IEEE Trans. Signal Process.*, vol. 62(24), pp.6425–6437, 2014.
27. B. Girault, "Stationary graph signals using an isometric graph translation," *European Signal Processing Conference (EUSIPCO)*, pp. 1516-1520, 2015.
28. L. Stanković, "A measure of some time-frequency distributions concentration," *Signal Processing*, Vol.81, No.3, pp.621-631, March 2001.
29. T. Thayaparan, LJ. Stanković, M. Daković, and V. Popović-Bugarin, "Micro-Doppler parameter estimation from a fraction of the period," *IET Signal Processing*, Vol. 4, No. 3, pp. 201–212, June 2010.
30. M. Daković, L. Stanković, and E. Sejdić, "Local Graph Signal Smoothness and Vertex-Frequency Representations", *submitted to Mathematical Problems in Engineering,* 2018

31. L. Stanković, M. Daković, and E. Sejdić, "Vertex-Frequency energy distributions" *IEEE Signal Processing Letters*, Vol. 25, no. 3, pp. 358–362, March 2018.
32. L. Stanković, M. Daković, and E. Sejdić, "Reduced Interference Vertex-Frequency Distributions" *IEEE Signal Processing Letters*, Vol. 25, no. 9, pp. 1393–1397, Sept. 2018.